

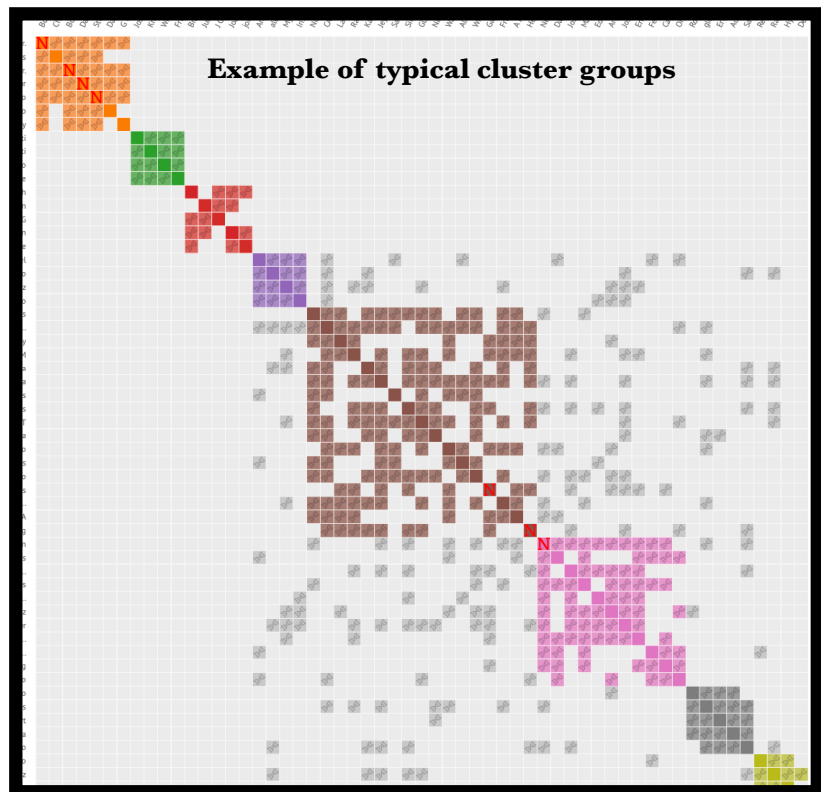
DNA Clustering: Automating Shared Matching

Alec Ferretti

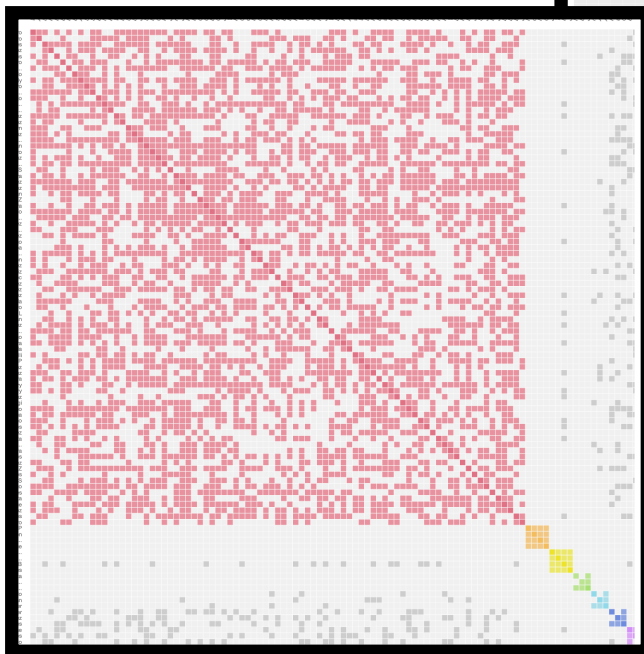
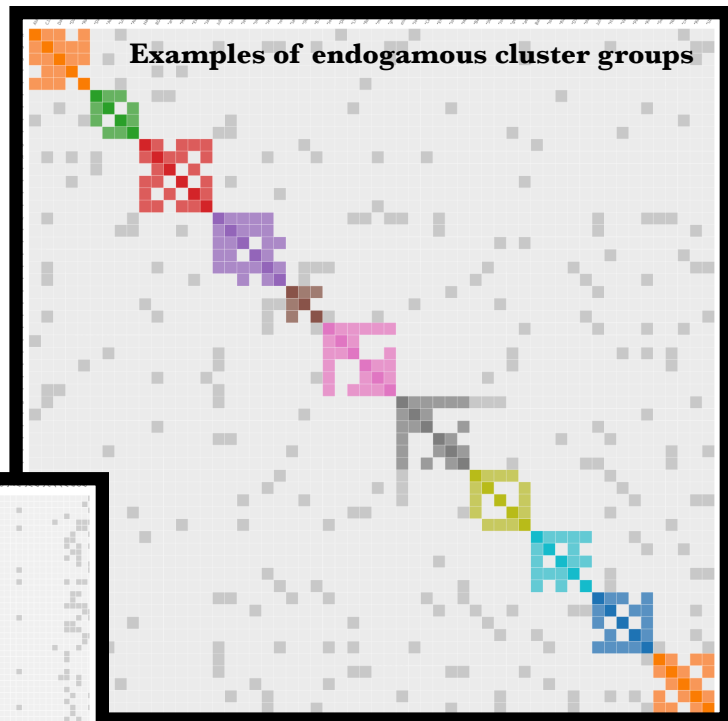
AncestorArchivist@gmail.com

Auto-Clustering

- Scary sounding term for a very handy tool
- Auto-clustering mechanizes the standard method of analyzing DNA matches
- Special software creates a easy to read chart of our shared matches
- Our matches who match each other are placed into clusters
- Each cluster likely descends from a common ancestor or otherwise has some genetic link
- This data is the same as what is provided from the testing company, but more concise
- Different peoples' clusters will look different depending on their genetic background
- Issues such as pedigree collapse and endogamy can result in less useful clusters
- People who have very distinct ethnic backgrounds will have very clean clusters
- The charts themselves are not so much useful as visually appealing
- However beneath each chart appears a table which shows the match data organized by cluster



- If one or two matches in a cluster have a known genetic connection to the tester, it can be postulated that other unknown matches from that cluster likely share the same genetic connection
- For testers who matches do not fall into neat clusters, it is possible to manipulate the data to reduce false positives by increasing thresholds



- It is also possible to analyze specific shared segments amongst members of cluster in order to look for genetic links
- The techniques of shared matching that all genealogists use have not changed: simply use auto-clusters to look for new leads amongst a sea of matches!

Referenced Websites:

Genetic Affairs: <https://geneticaffairs.com/>

Shared cM Project: <https://thegeneticgenealogist.com/2020/03/27/version-4-0-march-2020-update-to-the-shared-cm-project/>

DNA Painter: <https://dnainter.com/tools/sharedcmv4>